Épsilon - *Revista de Educación Matemática* *2017, Vol. 34, nº 95, 109-114*

La definición frecuentista de probabilidad a través de la simulación con el lenguaje de programación R

José Carlos Casas
Universidad de Córdoba
Carmen León-Mantero
Universidad de Córdoba
Noelia Jiménez-Fanjul
Universidad de Córdoba

Resumen: *Presentamos una secuencia de actividades mediante simulación a través del lenguaje de programación R, para mostrar a los maestros de Educación Primaria que el cálculo teórico de probabilidades se verifica cuanto mayor sea el número de ensayos realizados del experimento. Se refuerza además la importancia de que los elementos del espacio muestral de un experimento sean equiprobables para el correcto uso de la Regla de Laplace.*
Palabras clave: *Definición frecuentista de probabilidad, simulación con R, equiprobabilidad, maestros en formación.*

The frequentist definition of probability using R programming language simulation

Abstract: *We present a sequence of classroom activities for primary school teachers to work out probability using R programming language simulation. It is shown that for a given experiment the larger the number of trials performed, the closer the frequentist probability is to the expected value. It is also reinforced that the elements of the experiment sample space must be equally likely so as to use Laplace's rule appropriately.*
Keywords: *Frequentist definition of probability, computer simulation using R, equiprobability, teacher training.*

INTRODUCCIÓN

Las recientes disposiciones curriculares en relación a la asignatura de Matemáticas buscan que los alumnos, tras finalizar la Educación Primaria, sean capaces de describir, analizar, encontrar regularidades y hacer predicciones de situaciones que acontecen en la vida cotidiana, que les proporcione información para tomar decisiones y les permita abordarlas de manera adecuada.

Para ello es esencial que los maestros en formación dominen todos los contenidos matemáticos que en un futuro tendrán que enseñar en sus aulas a sus alumnos y, así, poder ejercer su labor docente con calidad.

El Real Decreto 126/2014 [RD 126/2014] (2014) organiza el currículo básico de la Educación primaria correspondiente a la asignatura de Matemáticas, en cinco grandes bloques. El último bloque, titulado *estadística y probabilidad*, incluye los contenidos:

- Carácter aleatorio de algunas experiencias.
- Iniciación intuitiva al cálculo de la probabilidad de un suceso.

Por otro lado, se establece el siguiente criterio de evaluación y sus correspondientes estándares evaluables, que definen los resultados de aprendizaje que los alumnos deben alcanzar con respecto a los contenidos anteriores:

- Observar y constatar que hay sucesos imposibles, sucesos que con casi toda seguridad se producen, o que se repiten, siendo más o menos probable esta repetición.
    - Identifica situaciones de carácter aleatorio.
    - Realiza conjeturas y estimaciones sobre algunos juegos (monedas, dados, cartas, lotería…).

Sin embargo, diversas investigaciones muestran que alumnos de todos los niveles educativos, presentan dificultades y obstáculos en relación a los contenidos anteriores. Serrado, Cardeñoso y Azcárate (2005) señalan que es común en los alumnos suponer que el resultado teórico de la probabilidad de un suceso va a cumplirse también en un número pequeño o limitado de pruebas, lo que podría estar fomentado porque en el aula se analice exclusivamente sucesos aleatorios de manera teórica.

Cuando lanzamos una moneda al aire, los resultados que se pueden obtener son CARA o CRUZ. Si la moneda está equilibrada, decimos que la probabilidad de que se obtenga CARA en el lanzamiento es 1/2 ya que es uno de los dos posibles resultados equiprobables. Esto puede entenderse de la siguiente forma: *Cada dos lanzamientos que realizamos, uno de ellos será CARA*. No obstante, esto no es cierto, ya que, al intervenir el azar, puede ocurrir que, tras 10 lanzamientos, sólo en uno de ellos se obtenga CARA, pero esto no significa que la probabilidad de obtener CARA sea 1/10.

Por otro lado, se ha observado en los alumnos la dificultad de establecer un espacio muestral que contenga los sucesos equiprobables de todo experimento aleatorio, lo que les lleva a calcular la probabilidad de manera incorrecta (Lecoutre, 1992).

Por todo lo anterior, esta propuesta presenta diversas actividades destinadas a mostrar a los alumnos del grado de Educación Primaria, que las probabilidades calculadas teóricamente con la regla de Laplace ocurren cuando el número de veces que se realiza el experimento aumenta considerablemente, así como, la necesidad de construir el espacio muestral de un experimento, de forma que, los elementos de éste sean equiprobables.

Batanero (2000) indica que la simulación de un experimento aleatorio en el aula, con un número elevado de ensayos, “permite condensar el experimento en el tiempo y en el espacio y operar con el experimento simulado para obtener conclusiones válidas para el experimento original" (p. 11)

Por ello, calcularemos las probabilidades de cada uno de los sucesos elementales de un experimento, mediante la regla de Laplace, y posteriormente simularemos la realización del experimento, un número elevado de veces que muestre que esas probabilidades se acercan a las calculadas teóricamente.

Para realizar la simulación del experimento se utilizará, a nivel muy básico, el software de libre distribución R (Versión 3.3.2; 2016). Éste puede ser descargado desde la web del proyecto: https://cran.r-project.org/. R es un potente lenguaje de programación, especialmente pensado para la realización de estudios estadísticos y el cálculo de probabilidades, y ampliamente integrado en la comunidad universitaria, que tiene, aun hoy, una escasa implantación en secundaria.

La gran desventaja con la que partimos al utilizar este software es la dificultad propia que tiene el desconocimiento del lenguaje de programación utilizado. No obstante, la reducida cantidad de instrucciones necesaria, disminuye la dificultad significativamente e introduce al alumno en el uso de lenguajes de programación a un nivel muy básico.

ACTIVIDADES

Actividad 1

Empezaremos en primer lugar por uno de los experimentos aleatorios más sencillos: el lanzamiento de una moneda equilibrada.

Se establecen que los dos resultados posibles que pueden obtenerse son CARA y CRUZ, escribiendo el espacio muestral como:

$$E=\{Cara,\ Cruz\}$$

y que, al ser la moneda equilibrada, la probabilidad de que ocurran cada uno de los sucesos es $1/2$.

A continuación, realizaremos la simulación del lanzamiento de la moneda con R:

Abrimos el programa, y creamos un nuevo script con la opción correspondiente del menú Archivo. A continuación, escribiremos en la ventana del script el comando que creará el vector que contiene a los elementos del espacio muestral. Este es:

Moneda=c("Cara", "Cruz")

Es importante indicar que el lenguaje hace distinción entre mayúsculas y minúsculas. Para ejecutar el comando y crear la variable Moneda basta con situar el curso sobre la línea de éste, hacer clic con el botón derecho y seleccionar la opción "correr línea o seleccionar". También con la combinación de teclas Ctrl + r.

Una vez creado, podemos comprobar que se ha generado correctamente, escribiendo el nombre de la variable y ejecutando dicha orden.

A continuación, generaremos los lanzamientos de la moneda con el comando *sample*. Su sintaxis es la siguiente:

sample(*Espacio muestral*, *nº de repeticiones*, *valor lógico*)

El valor lógico (TRUE o FALSE) indicará si en las repeticiones los elementos del espacio muestral pueden repetirse (con reemplazo) o no (sin reemplazo).

Para contar el número de veces que ha ocurrido cada resultado aplicaremos el comando *table*. Si además dividimos el resultado de éste por el número de veces que se ha repetido el experimento, éste nos devolverá la probabilidad obtenida en cada resultado.

Los comandos necesarios para la realización de 10, 100, 1000 y 10.000 lanzamientos se muestran en la Figura 1, junto con los resultados obtenidos.

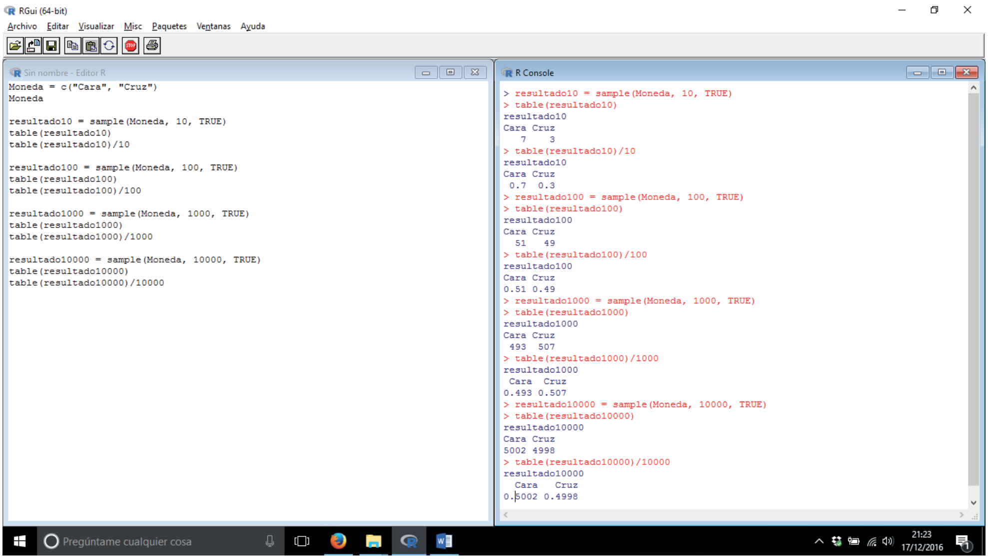

Figura 1. Resultado de la simulación del lanzamiento de una moneda.

Como puede observarse, cuando se lanza la moneda 10 veces, la cara se ha obtenido en 3 ocasiones y la cruz en 7, lo que representan probabilidades de 0.3 y 0.7 respectivamente. No obstante, cuando se aumenta el número de lanzamientos a 100, estas probabilidades se acercan hasta 0.51 para cara y 0.49 para cruz. Al aumentar a 10.000 el número de lanzamientos, las probabilidades son de 0.5002 y 0.4998 respectivamente.

Actividad 2

En el experimento consistente en lanzar 2 monedas, al alumno, se le presenta la disyuntiva entre estos dos espacios muestrales:

$$E = \{(Cara, Cara), (Cara, Cruz), (Cruz, Cruz)\}$$
$$E = \{(Cara, Cara), (Cara, Cruz), (Cruz, Cara), (Cruz, Cruz)\}$$

Es decir, ¿Es necesario especificar los distintos órdenes que pueden darse cuando los resultados son diferentes?

En el primero de los casos, la probabilidad de cada suceso elemental es 1/3, mientras que, en el segundo caso, esta probabilidad disminuye hasta 1/4, de modo que la probabilidad de obtener resultados diferentes es 1/2.

Para la simulación, ejecutaremos dos comandos *sample*, y los uniremos con el comando *paste*, que sirve para unir las componentes de los vectores de resultados.

Los resultados para esta nueva simulación se muestran en la Figura 2.

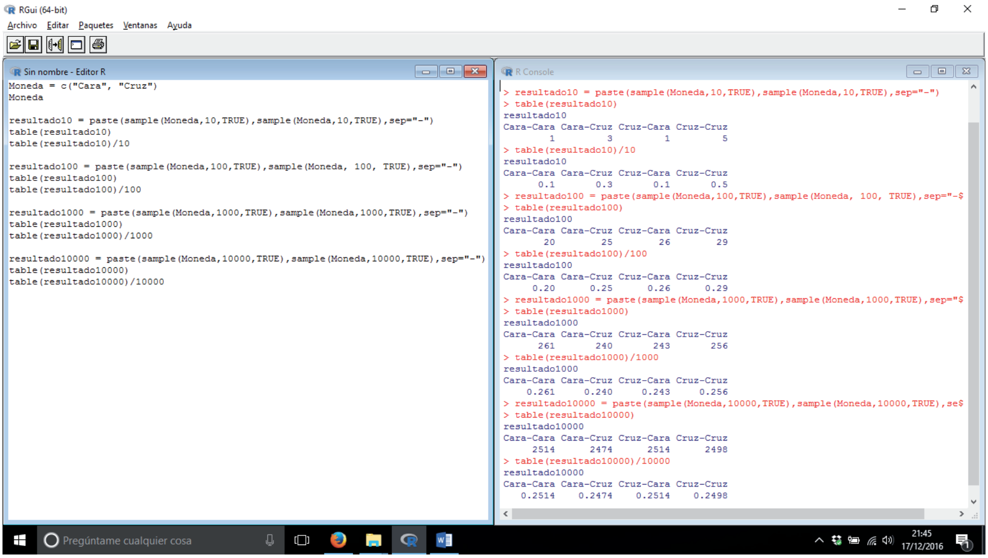

Figura 2. Resultado de la simulación del lanzamiento de dos monedas.

Como se observa en los resultados las probabilidades se aproximan, cuando aumentamos el número de repeticiones, a 0.25, por lo que el espacio muestral correcto es el segundo, si se quiere expresar éste con elementos equiprobables.

Este experimento puede repetirse para el lanzamiento de 3 o más monedas, añadiendo términos dentro del comando paste, separados por comas. También puede modificarse el número de repeticiones como se desee, para adaptarlo a otros ejemplos.

Actividad 3

Veamos ahora el experimento consistente en lanzar un dado. En este caso, repetiremos el lanzamiento de éste 60, 600, 6.000 y 60.000 veces, y analizaremos el número de veces que ha sucedido cada uno de los seis resultados posibles.

El espacio muestral del experimento es:

$$E = \{1,2,3,4,5,6\}$$

Y si el dado es equilibrado, la probabilidad de cada uno de los sucesos elementales es $1/2 = 0.1\overline{6}$.

La Figura 3, muestra las simulaciones realizadas en este caso.

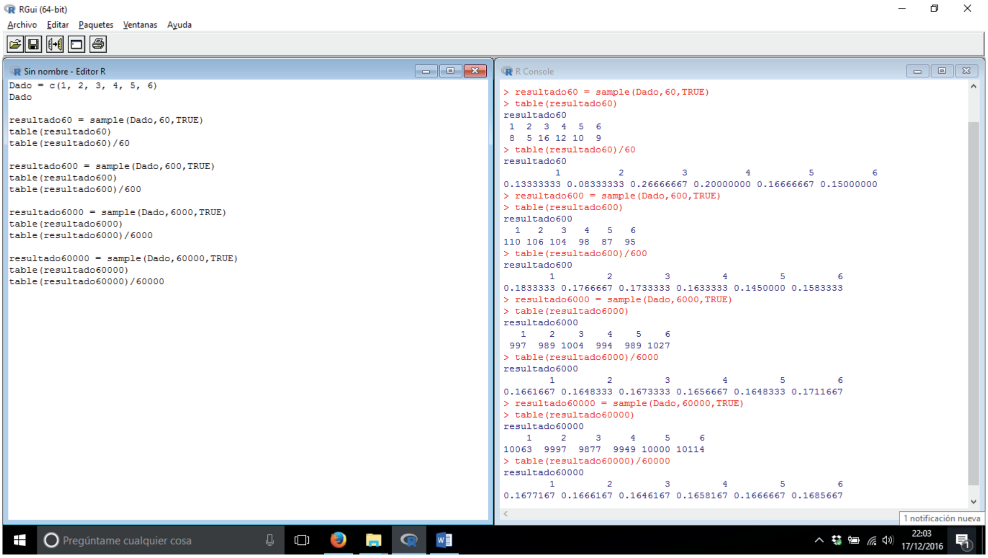

Figura 3. Resultado de la simulación del lanzamiento de un dado.

Cuando se repite el lanzamiento 60 veces, el único resultado con una probabilidad igual a la teórica es el 5, que se ha obtenido 10 veces. Cuando se repite 60.000 veces, la proporción de veces que se obtiene cada uno de los resultados difiere de la teórica en menos de dos milésimas.

Actividad 4

Por último, analizaremos el experimento consistente en el lanzamiento de dos dados y la suma de los resultados.

En este caso, el espacio muestral que en primer lugar viene a la mente del alumno es:

$$E = \{2,3,4,5,6,7,8,9,10,11,12\}$$

No obstante, los elementos que lo forman no son equiprobables, lo que obliga a expresarlo de la siguiente forma:

$$E = \left\{ \begin{array}{l} (1,1),(1,2),(1,3),(1,4),(1,5),(1,6), \\ (2,1),(2,2),(2,3),(2,4),(2,5),(2,6), \\ (3,1),(3,2),(3,3),(3,4),(3,5),(3,6), \\ (4,1),(4,2),(4,3),(4,4),(4,5),(4,6), \\ (5,1),(5,2),(5,3),(5,4),(5,5),(5,6), \\ (6,1),(6,2),(6,3),(6,4),(6,5),(6,6) \end{array} \right\}$$

Por lo que las probabilidades son:

Suma	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Probabilidad (Fracción)	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$
Probabilidad (Decimal)	$0.02\overline{7}$	$0.0\overline{5}$	$0.08\overline{3}$	$0.\overline{1}$	$0.13\overline{8}$	$0.1\overline{6}$	$0.13\overline{8}$	$0.\overline{1}$	$0.08\overline{3}$	$0.0\overline{5}$	$0.02\overline{7}$

En este caso, basta con sumar dos simulaciones del lanzamiento de un dado, como el realizado anteriormente. Los resultados para 36, 360 y 3.600 lanzamientos de dos dados se muestran en la Figura 4. En ésta se puede observar cómo las probabilidades se aproximan, también en esta ocasión, a las teóricas.

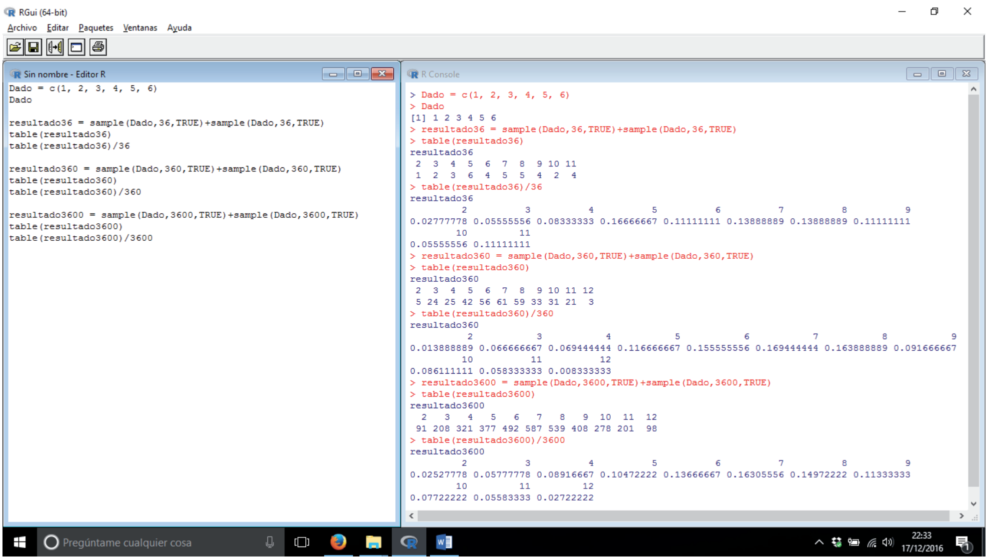

Figura 4. Resultado de la simulación del lanzamiento de dos dados y su suma.

Los resultados de estas simulaciones pueden también representarse mediante un diagrama de barras con el comando *barplot*. Este gráfico para el caso de las 3600 repeticiones, cuyo comando completo es:

$$barplot(table(resultado3600)/3600)$$

El gráfico obtenido se muestra en la Figura 5, donde puede observarse que la mayor probabilidad se concentra en los valores centrales por el mayor número de resultados que contienen.

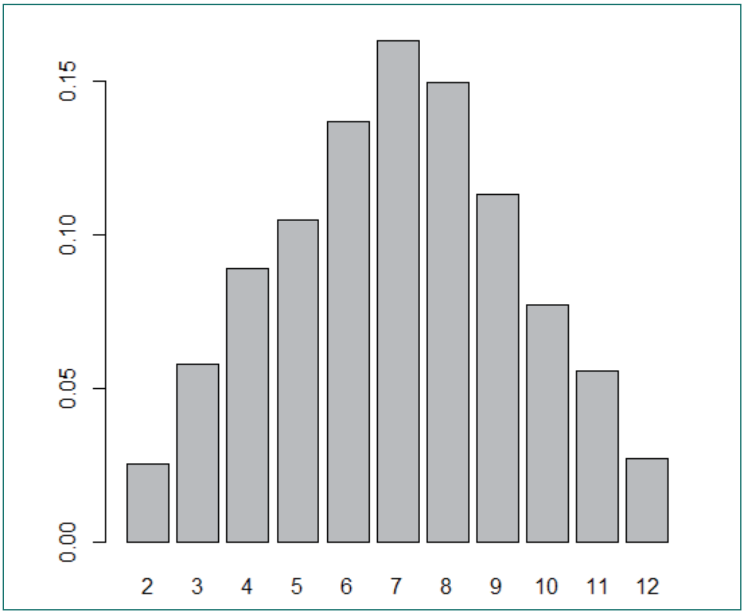

Figura 5. Gráfico de barras del resultado obtenido al lanzar 3600 veces un par de dados.

CONCLUSIONES

Como puede observarse, con unos sencillos comandos, puede simularse, en R, un gran número de experimentos aleatorios para analizar de forma experimental la probabilidad de cada suceso elemental y compararla con la probabilidad teórica de cada suceso. Esto puede ayudar, además, a comprender dos conceptos fundamentales como son la definición frecuentista de probabilidad y la necesaria equiprobabilidad en el espacio muestral para la aplicación de la regla de Laplace.

REFERENCIAS

Batanero, C. (2000). ¿Hacia dónde va la educación estadística? *Blaix, 15*(2), 13.

Lecoutre, M. P. (1992). Cognitive models and problem spaces in “purely random” situations. *Educational studies in mathematics, 23*(6), 557-568.

Real Decreto 126/2014, de 28 de febrero, por el que se establece el currículo básico de la Educación primaria (2014). BOE, 52, 19349-19420.

Serrado, A., Cardeñoso, J.M. y Azcárate, P. (2005). Los obstáculos en el aprendizaje del conocimiento probabilístico: su incidencia desde los libros de texto. *Statistics Education Research Journal, 4*(2), 59-81.